

2
RS

jc997 U.S. PTO
10/042351
01/11/02

501.41076X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): TAKASE, et al.
Serial No.: Not assigned
Filed: January 11, 2002
Title: PACKET SWITCHING SYSTEM
Group: Not assigned

LETTER CLAIMING RIGHT OF PRIORITY

Honorable Commissioner of
Patents and Trademarks
Washington, D.C. 20231

January 11, 2002

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on Japanese Application No. (s) 2001-047548 filed February 23, 2001.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP

Carl I. Brundidge
Registration No. 29,621

CIB/amr
Attachment
(703) 312-6600

日本国特許庁 2.

JAPAN PATENT OFFICE

2-20-02

3100-2289

RS WN

JG997 U.S. PRO
10/042351
01/11/02

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office

出願年月日
Date of Application:

2001年 2月 23日

出願番号
Application Number:

特願2001-047548

出願人
Applicant(s):

株式会社日立製作所

CERTIFIED COPY OF
PRIORITY DOCUMENT

2001年10月19日

特許庁長官
Commissioner,
Japan Patent Office

及川耕三



出証番号 出証特2001-3092536

【書類名】 特許願

【整理番号】 H00022891A

【あて先】 特許庁長官 殿

【国際特許分類】 H04L 12/56

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所中央研究所内

【氏名】 高瀬 誠由

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所中央研究所内

【氏名】 豊田 英弘

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所中央研究所内

【氏名】 森脇 紀彦

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社 日立製作所

【代理人】

【識別番号】 100075096

【弁理士】

【氏名又は名称】 作田 康夫

【電話番号】 03-3212-1111

【手数料の表示】

【予納台帳番号】 013088

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

特2001-047548

【物件名】 図面 1

【物件名】 要約書 1

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 パケット処理装置

【特許請求の範囲】

【請求項 1】

パケット処理装置であって、
複数の入力回線処理部と、
それぞれ、上記複数の入力回線処理部に接続される複数の入力バッファ部と、
複数の出力回線処理部と、
上記複数の入力バッファ部と、上記複数の出力回線処理部とに接続されるクロスバスイッチとを有し、
上記複数の入力バッファ部の各入力バッファ部は、上記複数の出力回線処理部に
に対応して設けられた複数のキューバッファを有し、
上記複数の入力バッファ部の全てのキューバッファの中から、何れのキューバッファ
に対して、上記クロスバスイッチへパケットを送出する権利を付与するか
のアービトレーションは、上記全てのキューバッファの各キューバッファについて
計算される、そのキューバッファから上記クロスバスイッチにパケットが送出
される時間間隔に関する情報と、そのキューバッファのキュー長に関する情報と
に基づき行われるパケット処理装置。

【請求項 2】

請求項 1 に記載のパケット処理装置であって、
上記全てのキューバッファの各キューバッファに対応に、そのキューバッファから
パケットが上記クロスバスイッチに送出される時間間隔を測定する手段と、
上記複数の入力バッファ部の各キューバッファに対応に、そのキューバッファの
キュー長を測定する手段とを更に有するパケット処理装置。

【請求項 3】

請求項 1 又は請求項 2 の何れかに記載のパケット処理装置であって、
上記全てのキューバッファの各キューバッファからパケットがあふれ出ない
ようにするため、上記アービトレーションは、上記時間間隔に関する情報よりも
上記キュー長に関する情報にウエイトを置いて行われるパケット処理装置。

【請求項4】

請求項1又は請求項2の何れかに記載のパケット処理装置であって、
上記全てのキューバッファの各キューバッファにパケットが滞留する時間を
短くするために、上記アビトレーションは、上記キー長に関する情報よりも
、上記時間間隔に関する情報にウェイトを置いて行われるパケット処理装置。

【請求項5】

パケット処理装置であって、
複数の入力回線処理部と、
複数の出力回線処理部と、
それぞれ、上記複数の出力回線処理部に対応して設けられた複数のキューバッ
ファを有するとともに、上記複数の入力回線処理部に接続される複数の入力バッ
ファ部と、
上記複数の入力バッファ部と、上記複数の出力回線処理部とに接続されるクロ
スバスイッチと、
上記複数の入力バッファ部の全てのキューバッファの中から、何れのキューバ
ッファに対して、上記クロスバスイッチへパケットを送出する権利を付与するか
のアビトレーションを行うアビタと、
上記複数の入力バッファ部の全てのキューバッファに対して、それらのキュ
ーバッファから上記クロスバスイッチにパケットが送出される時間間隔に関する情
報と、それらのキューバッファのキー長に関する情報とに基づき、上記クロス
バスイッチへパケットを送出する権利の取得に関する優先度情報を求める手段と
を有し、
上記アビタは、上記複数の入力バッファ部の全てのキューバッファに対して
求められた上記優先度情報に基づき、上記アビトレーションを行うパケット処
理装置。

【請求項6】

請求項5に記載のパケット処理装置であって、
上記全てのキューバッファの各キューバッファ対応に、そのキューバッファか
らパケットが上記クロスバスイッチに送出される時間間隔を測定する手段と、

上記複数の入力バッファ部の各キューバッファ対応に、そのキューバッファのキュー長を測定する手段とを更に有するパケット処理装置。

【請求項7】

請求項5又は請求項6の何れかに記載のパケット処理装置であって、
上記全てのキューババッファの各キューバッファからパケットがあふれ出ない
ようするため、上記優先度情報は、上記時間間隔に関する情報よりも上記キュ
ー長に関する情報にウェイトを置いて求められるパケット処理装置。

【請求項8】

請求項5又は請求項6の何れかに記載のパケット処理装置であって、
上記全てのキューババッファの各キューバッファにパケットが滞留する時間を
短くするために、上記優先度情報は、上記キュー長に関する情報よりも、上記時
間間隔に関する情報にウェイトを置いて求められるパケット処理装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、入力バッファ型クロスバスイッチのパケット処理装置に関し、特に
、当該パケット処理装置のアービタ方式に関する。

【0002】

【従来の技術】

入力回線毎に单一FIFOを持つ入力バッファ型クロスバスイッチのパケット
処理装置が知られている。このパケット処理装置では、複数の入力回線から入力
された複数のパケットが特定の出力方路に集中すると、HOL(Head Of
Line)ブロッキングが発生する。このため、スループットが58.6%まで
しか上がらないことは良く知られている。そこで、入力バッファ部に、出力方路
毎にVoQ(Virtual Output Queue)を設けるものが知られて
いる。

【0003】

また、可変長パケットのスイッチング方式として、可変長パケットをある一定
の固定長の短いセルに分割し、1セル単位でアービトレーションを行なうものと

、特開2000-232482号公報に示されているように、可変長パケットをある一定の大きな固定長の入れ物に複数個詰めこみ、この入れ物単位でアービトレーションを行なうものとがある。このように、一回のアービトレーション処理でスイッチへ送出されるデータは、固定長の短いセルである場合、固定長の長い入れ物である場合があるが、本明細書では、一回のアービトレーション処理でスイッチに送出されるデータを1セグメントと呼ぶことにする。

【0004】

V○Qを持つ入出力バッファ型クロスバスイッチでは、クロスバスイッチ部に通常バッファリング要素を持たないため、複数のV○Qからの出力が競合しないように、アービトレーション機能を設ける必要がある。アービトレーションの良し悪しは、スイッチのスループットに大きな影響を与える。

【0005】

従来のアービトレーションの方法として、次の3つの方法が提案されている。第1に、V○Qのパケットの有無を見てラウンドロビンで送出キューを選択する方法。この方法については、若山浩二 他、"大容量パケットスイッチ構成法に関する一検討"、信学技報IN98-102、及び特開2000-78148号公報に開示されている。第2に、時間をパラメーターとして見て、送出キューを選択する方法。この方法については、須藤俊之 他、"入力バッファスイッチのスケジューリングの一検討と試作"、信学技報SSE99-118に開示されている。第3に、キュー長を見て送出キューを選択する方法。この方法については、大場義洋、"キュー長を用いた重み付け公平パケットスケジューリングアルゴリズムの提案および性能評価"、信学技報SSE96-56に開示されている。

【0006】

【発明が解決しようとする課題】

上記の従来の技術では、スイッチに不均衡負荷が掛かった時に生じる問題点を図16を用いて説明する。図16は、4×4スイッチの入力ポート30の概念図である。31-1,-2,-3,-4はそれぞれの出力方路行きのV○Qである。各V○Q内の四角形はセグメントを表している。31-1には他のキューと比べて高負荷なトラヒックが掛かっており、31-4には他のキューと比べて低負荷

なトラヒックが掛かっている。

【0007】

VoQのパケットの有無を見てラウンドロビンで送出キューを選択する方法では、不均衡負荷下の状態において、ラウンドロビンにより公平に全VoQからパケットが読み出されるため、低負荷キューは高負荷キューの影響を全く受けることなくパケットを送出することが出来る。一方、高負荷なトラヒックが掛かっているVoQは、キュー長が伸びるため、遅延時間が他のVoQに比べ大きくなる。そのため、高負荷キューのバッファVoQ1がオーバーフローを起こし、図に示すように、セグメント32Aが廃棄される可能性がある。

【0008】

アービトレーションでVoQにGrant（送出権）を与える一つのパラメーターとしてパケットの送出間隔を用いる方式でも、同じような現象がおこる。高負荷が掛かっているVoQではキュー長が長くなる。キュー長が長くなると、パケットのVoQでの待ち時間が長くなる。そのため、待ち時間の長いパケットを有するVoQから優先的にパケットを送出する方式を採用すると、パケットは高負荷キューから優先的に送出されることになる。しかし、すべてのパケットについて待ち時間を管理するとなると、大量のカウンタが必要となり、実現は非常に困難にである。このため、多くの場合パケットがキューの先頭に到着した時から待ち時間をカウントする方式を採用している。これは、一アービトレーション周期毎にパケットが送出されなかったらその都度カウンターを1づつ増やし、パケットが送出されたらそのVoQのカウンターを0に戻すことである。この方式だと、高負荷が掛かっているVoQでも、一度パケットを送出してしまうと、カウンターが0に戻されるため、他の低負荷キューと同じ条件になる。このため、結局は不均衡負荷下において高負荷キューの遅延時間が大きくなり、バッファのオーバーフロー、パケットの廃棄を引き起こす可能性がある。

【0009】

キュー長を見てアービトレーションを行なう方法では、同じような不均衡負荷が掛かった時には高負荷キューから効率良くパケットが読み出される。そのため、高負荷キューの遅延時間は小さくなり、バッファのオーバーフローも防ぐこと

が出来る。しかし、31-4の低負荷キューに送出権が巡ってこなくなる現象（この現象はstarvationと呼ばれる）が発生するなどの問題がある。

【0010】

したがって、従来のアービタでは遅延に厳しい音声パケットや、廃棄に厳しい重要なデータパケットなどの品質劣化が起こる可能性がある。

【0011】

【課題を解決するための手段】

本発明のパケット処理装置は、V○Qのパケット送出間隔と、キュー長とに基づき、クロスバスイッチに対するパケットの送出権を何れのV○Qに与えるかを決定する。

【0012】

本発明のパケット処理装置の一実施例においては、入力回線毎に、パケット各V○Qに貯まっているセグメントの量を管理するキュー情報管理部と、各V○Qのセグメントの送出間隔を管理する送出間隔管理部と、それら二つの情報から各V○Qにレベルを割り振るアービタリクエスト（ARB-REQ）生成部とを設ける。アービタは、各V○Qに割り振られたレベルに基づき、何れのV○QにGrantを与えるかを決定する。また、一実施例においては、セグメントの送出間隔を、V○Qのキュー長に優先させて、V○Qのレベルを設定する手段、または、これとは逆に、V○Qのキュー長を、セグメントの送出間隔に優先させて、V○Qのレベルを設定する手段を設ける。

【0013】

【発明の実施の形態】

図1は、本発明のパケット処理装置の一実施例を示す図である。本実施例のパケット処理装置では、V○Q11とクロスバスイッチ19とを結ぶ主信号線と異なる別線18により、ARB-REQ情報をアービタに送る。

【0014】

入力回線処理部16-i ($i = 1, \dots, n$) は、入力回線103-iから入力されたパケットのパケットヘッダーを解析することにより、そのパケットの宛先情報を抽出する。入力バッファ10-iは、出力方路に対応にn個のV○Qキュー

—11—iを有している。入力回線処理部16は、V○Q制御部12に、その抽出した宛先情報を渡す。V○Q制御部12は、そのパケットを、宛先情報に応じた出力方路に対応するV○Qに書き込むように、入力バッファ10に指示を与える。このようにして、パケットは指示されたV○Q11に書き込まれる。

【0015】

V○Q制御部12は、V○Q11毎に、滞留しているセグメントの量と、送出間隔との情報を管理する。ARB-REQ生成部13は、それらの情報に基づき、各V○Qにレベルを割り振る。各V○Qのレベルは、信号線18により、アビトレーション周期でアビタ14に集められる。アビタ14は、その情報から、どのV○QにGrantを与えるかを決定する。Grant情報は、信号線15により、ARB-ACKとして、各V○Q制御部12-i (i=1, ..., n)に伝えられ、同時にアビタは、そのアビトレーション結果をクロスバスイッチ19内の経路構成に反映する。V○Q制御部12は、そのARB-ACK情報に基づき、入力バッファ10に対して、どのV○Qからセグメントを送出するかを伝える。入力バッファ10より送出されたセグメントは、クロスバスイッチ19によりスイッチングされ、出力回線処理部17に送信される。出力回線処理部17は、クロスバスイッチ19から受信したセグメントからパケットを再構成し、それを出力回線104に送出する。

【0016】

図2を参照して、V○Q制御部12とARB-REQ生成部13の構成を詳述する。入力回線処理部16内のパケットヘッダー解析部401で処理されたパケットヘッダーは、V○Q制御部12のライトアドレス(WA)生成部403に送られる。WA生成部403では入力されたパケットをどのV○Q11に書きこむかを管理する。WA生成部403は、WA制御信号線412により、メモリアドレス情報を入力バッファ10に送信し、パケットを、その宛先に対応したV○Q11にパケットを書きこむよう指示を与える。同時に、WA生成部403は、キュー長管理部405と送出間隔管理部406に対しても、入力バッファ10に書き込まれたパケットの情報を伝える。キュー長管理部405は、入力バッファ10内の複数のV○Qに対応し、キュー長カウンタ410を有している。図2に示す

例では、入力バッファ10に4つのV○Qがある場合を示しているため、キュー長管理部は、4つのキュー長カウンタ410を有している。キュー長管理部405は、現在のキュー長に対して、今回入ってきたパケットのセグメント長分だけキュー長カウンター410の数値をインクリメントする。送出間隔管理部406は、入力バッファ10内の複数のV○Qに対応し、送出間隔カウンター411を有している。送出間隔管理部406は、パケットが入力されたV○Qに、パケットが滞留している場合には何もしない。そうでない場合には、送出間隔管理部406は、一アビトリレーション周期毎に数値を1ずつインクリメントするように、そのV○Qに対応する送出間隔カウンター411に指示を与え、送出間隔時間を管理する。言い換えれば、送出間隔カウンター411が示すデータは、どの位の時間、対応するV○Qからパケットが送出されていないかを示すものである。

【0017】

リードアドレス（RA）生成部404は、信号線15により送信してきたARB-ACK情報に基づき、信号線413により、入力バッファに対して、どのV○Qからセルを送出するかを伝える。同時に、リードアドレス（RA）生成部404は、キュー長管理部405と送出間隔管理部406にも入力バッファ10から読み出されるパケットの情報を伝える。キュー長管理部405は、クロスバスイッチ19にパケットを送出したV○Qに対応するキュー長カウンター410をディクリメントする。また、送出間隔管理部406は、そのV○Qに対応する送出間隔カウンター411の値を0に戻す。

【0018】

キュー長管理部405と送出間隔管理部406における情報は、信号線414により、ARB-REQ生成部13に送信される。ARB-REQ生成部13は、入力バッファ10内の各キューに対応し、ARB-REQ生成パート409を有する。各ARB-REQ生成パート409-1~-4は、ARB-REQ生成部13から受信した情報に基づき、対応するキューのレベル付けを行う。レベル付けを行う際は、レベル付与マトリクス416を参照する。レベル付与マトリクス416はマイコン20により、そのノードに入力されるトラヒック特性に合わせたアビタにユーザがチューニングすることが可能である。これらのレベルは

、信号線18により、アービタ14に送信される。

【0019】

図3は、レベル付与マトリクス416の一実施例を示す図である。レベル付与マトリクスは、横軸にセグメント送出間隔71、縦軸にV○Qに滞留しているセグメント数72をとっており、送出間隔時間が長くなるにつれ、そして、V○Qに滞留しているセグメント数が増えるにつれV○Qに付与されるレベルが大きくなるように作られている。レベル付与マトリクスは、キュー長(V○Qに貯まっているセグメント量)とセグメント送出間隔の二つのファクターから算出されるが、V○Qに入ってきたパケットが要求遅延時間以内に送出されるようにレベル付けを行うことも可能である。要求遅延時間とは、そのスイッチ内で許される最大の遅延時間のことである。

【0020】

高負荷の掛かっているV○Qは他の低負荷のV○Qに比べ、パケットの遅延時間が大きくなる傾向にある。そのため、キューにたまっているセグメント数が多いV○Qには大きなレベルを付け、優先的にGrantが与えられるようにする。良い判れば、このレベルは、Grant取得に関する優先度情報を表すものである。また、低負荷キューからも要求遅延時間内にはパケットを送出できるようにするため、送出間隔が長くなると、V○Qレベルが増加するようにしている。

【0021】

V○QレベルLは、例えば、次式により求める。

【0022】

【数1】

$$L = \frac{1}{\ln\left(\frac{M-a \cdot t}{b \cdot s} \times e\right)} \times 15$$

【0023】

ここで、M：タイムアウト、t：送出間隔、s：現在のセグメント数、a：送出間隔係数、b：キュー長係数である。

【0024】

$(M - a \cdot t) / b \cdot s \leq 1$ となった時にV○Qレベルは最大値を取る。V○Qレベルが最大値になった時、そのV○Qは、送出権を得る可能性が高い状態になったことを意味する。

【0025】

Mは、スイッチの要求遅延時間と一アビトレイション時間から決まる値であり、レベル付与マトリクスのセグメント送出間隔71の最大値と、V○Qに滞留するセグメント数72の最大値とを決める値となる。スイッチに求められている要求遅延時間をT、一アビトレイション時間をta、スイッチの入力ポート数をnとすると、Mは次式により求めることが出来る。

【0026】

【数2】

$$M = \frac{T}{t_a} - n$$

【0027】

要求遅延時間Tの間に何回アビトレイションが出来るかは T / t_a で求めることが出来る。全V○Qのレベルが同時に最大値になっていると仮定すると、送出権が与えられるまでに、最大nアビトレイション時間待たされるV○Qが現れる。したがって、パケットがそのV○Qに滞在する時間が $T / t_a - n$ になる時に、V○Qレベルが最大値をとるようにそのレベルを設定すると、一つしかセグメントが存在しないV○Qであっても、要求遅延時間以内にセグメントを送出することが出来るようになる。図3に示したレベル付与マトリクスからもわかる

ように、一つしかセグメントがないV○Qにおいても、送出間隔が大きくなるにつれてV○Qレベルは上がっていくので、必ずしも最大遅延量の要求時間までセグメントが送出されないことにはならない。

【0028】

図3に示したレベル付与マトリクスは、タイムアウトM=20, a=1, b=1に設定したものである。送出間隔係数a, キュー長係数bを変化させることで、送出間隔重視のアービトレーションやキュー長重視のアービトレーションに変えることが出来る。例えば、音声など遅延に厳しいデータを多く処理しなければならない場所に、本実施例のパケット処理装置を使用する時には、低負荷キューに遅延に厳しいデータがキューイングされた時、低負荷キューの遅延時間を出来る限り抑える設定に変更する。具体的には、送出間隔係数aの値を1よりも大きな値にすることで、セグメントがあまり貯まっていないV○Q付与されるレベルについても、短い送出間隔で大きなレベルを取るように設定することが出来る。図4は送出間隔係数aの値を2に設定した時のレベル付与マトリクスの状態を表している。図3のレベル付与マトリクスと比較すると、V○Qレベルは、セグメントの送出間隔が小さい時から大きな値となっている。このため、低負荷キューからも短い遅延時間でセグメントを送出することが出来る。

【0029】

逆に、遅延時間よりも、パケットの廃棄に厳しいデータを多く処理しなければならない場合に、本実施例のパケット処理装置を使用する場合、低負荷キューの遅延時間は若干延びるが、高負荷がかかり出したキューから優先的にセグメントを送出してバッファのオーバーフローを防止するようにする方が良い。このためには、キュー長係数bの値を1よりも大きな値にすることで、キュー長の変化に敏感に反応するV○Qレベル付けを行えば良い。高負荷が掛かり出すとキュー長は長くなりだす。図5はキュー長係数を2に設定した時のレベル付与マトリクスの状態を表している。図3のレベル付与マトリクスと比較すると、V○Qにあるセグメント数が少ない所でも大きなレベルを取っている。そのため、キュー長が長くなりだしたV○Qに対して、早めに大きなレベルを与え送出権を与えやすくすることでバッファのオーバーフローを事前に防ぐことが出来る。また、全く送

出間隔を考えず、キュー長のみを見てアービトレーションを行いたい時には、送出間隔係数 a の値を 0 に設定することで対応可能となる。

【0030】

数 1 により算出される $V \circ Q$ レベルは、すべて四捨五入され整数としてレベル付与マトリクスに表されている。また、 $M - a \cdot t < b \cdot s$ や $M \leq a \cdot t$ となつた時は $0 < L \leq 15$ 以外の値を取る。しかし、こういう状況になる時には、すでに $V \circ Q$ レベルが最大のレベル 15 を超えていることを表しているので、この数式により $0 < L \leq 15$ 以外の数値が与えられた $V \circ Q$ に対しては、レベル 15 を与える。

【0031】

各 A R B - R E Q 生成部から $V \circ Q$ 每のレベル情報がアービタ 14 に集められる。

【0032】

図 6 は、アービタ 14 の一実施例を示す図である。全 $V \circ Q$ レベル収集部 121において、全 $V \circ Q$ のレベル情報が集められる。

【0033】

図 7 は、ある時点における、Request のあった全 $V \circ Q$ からの状態を表している。マトリクスの列は出力回線番号 131 を表しており、行は入力回線番号 132 を表している。例えば、入力回線番号が 1 で出力回線番号が 1 の欄には、図 1 の入力バッファ 10-1 における、出力回線処理部 17-1 に送信されるパケットがキューイングされる $V \circ Q$ 11-1-1 に付与されたレベルが格納される。また、マトリクス中の \bigcirc 133 は $V \circ Q$ レベルが同じであり、空白部は “ \bigcirc ” よりも小さなレベルが付けられているとする。

【0034】

このような状態で、 $V \circ Q$ に最も効率良く Grant を与えるには、入力回線番号 2 で出力回線番号 4 の $V \circ Q$ と、入力回線番号 4 で出力回線番号 2 の $V \circ Q$ に Grant を与える必要がある。このアービトレーションによれば、少なくとも 2 つの $V \circ Q$ に送出権を与えることが出来る。一方、マトリクスの入力回線番号 2 で出力回線番号 2 の $V \circ Q$ に送出権を与えると、入力回線番号 2 で出

力回線番号4のV○Qと、入力回線番号4で出力回線番号2のV○Qには送出権を与えることが出来ない。したがって、この回のアビトリエーションでは最上位レベルのV○Qに1つしかGrantを与えることができなくなってしまう。効率良くGrantをV○Qに与えるために、入力別勝ち抜き処理部122-1は、同一入力回線番号に属するV○Qごとに、図8に示す勝ち抜き戦51を行ない、その中でV○Qレベルの一番高い（その入力回線番号の中では一番送出要請が高い）V○Qを選択する。

【0035】

図8において、数字は、同一の入力回線番号に属するV○Qのレベルを表している。図8は、一番レベルが高いV○Qが勝ち抜く様子を示している。図8では同じレベルのV○Qが2つ存在しているが、こういう時はどちらか一方を勝ち残らせるのではなく同一レベルのV○Qはすべて勝ち残らせるようとする。これは、図7を用いて説明したように、効率良くGrantを与えるためである。

【0036】

同様にして、出力別勝ち抜き処理部122-2は、同一出力回線番号に属するV○Q毎に、その中でV○Qレベルの一番高い（その入力回線番号の中では一番送出要請が高い）V○Qを選択する。

【0037】

勝ち抜き戦の結果、入力方向と出力方向において図9の勝敗組み合わせ141が考えられる。レベル再評価部124は、各V○Qのレベル情報を、図10に示すレベル再評価表61に則り、0~3の4レベルに縮退させ、再評価を行う。

【0038】

0~3の4レベルに再評価されたV○Qは、同一レベルV○Q選択部125において、レベル3のV○Qから順にピックアップされる。ここでピックアップされたV○QはGrant付与部126でラウンドロビン選択によってGrantが与えられる。

【0039】

Grantの与えられたV○Qと同一入力回線番号又は同一出力回線番号に存在するV○Qからは、このアビトリエーション周期において、これ以上Grant

tを与えることは出来ないので、送出権剥奪部127において、送出権を奪う。

【0040】

送出権を剥奪されたV○Qの情報は、信号線123により、全V○Qレベル収集部121に通知される。その情報から、まだ送出権の剥奪されていないV○Qの間だけで再度勝ち抜き戦を行ないレベルの再評価をする。その後、先程の説明と同様に同一レベルV○Q選択部125によってレベル3のV○Qがピックアップされ、Grant付与部によって、ラウンドロビン選択でGrantを与える。この反復作業を繰り返すことにより、最適の入力回線と出力回線との組み合わせを作り出すことが出来る。

【0041】

上記の工程でGrantを得たV○Qと同一入力回線番号又は同一出力回線番号に存在するV○Qからはこの回のアビトリレーションではこれ以上Grantを与えることは出来ないので、送出権剥奪部127は、そのV○Qから送出権を奪う。送出権を剥奪されたV○Qの情報は、信号線129により、同一レベル選択部125に通知される。同一レベルV○Q選択部125は、まだ送出権のあるレベル2のV○Qをピックアップし、Grant付与部によって、ラウンドロビン選択でGrantを与える。

【0042】

その後、レベル2のV○QにGrantを与えたのと同じ工程により、まだ送出権のあるレベル1のV○QにGrantが与えられる。その次に送出権があるレベル0のV○QにもGrantが与えられる。

【0043】

Grant情報はARB-ACK生成部128によりARB-ACK情報に変えられ、信号線18により入力バッファ管理部12のRA生成部404に送られる。RA生成部404は、ARB-ACK情報に基づき、セグメントを送出可能なV○Q11に対して、信号線413により、セグメントを送出するよう指示を送る。それと同時に、キュー長管理部405と送出間隔管理部406にも、どのV○Qからセグメントを送出すべきかを通知する。キュー長管理部405ではGrantの出たV○Qのセグメント数を管理しているキュー長カウンタ410の

値を1減らす。送出間隔管理部406ではGrantの出たV○Qの送出間隔を管理している送出間隔カウンタ411の値を0に戻す。

【0044】

図11は、上述したアービタ14における一連の処理を、フローチャートで表したものである。各入力バッファから集められたV○Qレベルは入力別、出力別で勝ち抜き戦が行なわれ81、各V○Qの送出要請の強い順にレベル3～0までのレベルで再評価され直す82-1～-4。まず、レベル3のV○Qを同一レベル選択部125でピックアップ83し、それらの物にラウンドロビンでGrantを与える83-1（このラウンドロビンは公平さを保たせるため、ポインターを持った2DRR（高城衛 他，“パケットの優先度を考慮した出力ポート別入力キュー付きパケットスイッチの改良” 信学技報SSE97-13）等を採用すれば良い）。Grantの与えられたV○Qと同じ入力、出力にあるV○Qからは同じアービトレーション周期に同時にセグメントを送出することは出来ないので、V○Qの送出権を剥奪83-2する。

【0045】

本実施例では、アービタの性能を高めるため、次にレベル2の物にGrantを与えるのではなく、Grantの出でていないV○Qのレベルを再評価前の0～15までのレベルに一旦戻し、もう一度勝ち抜き戦を行ない、レベルの再評価を行なう84。この行程を繰り返すほど、アービトレーションで選ばれるキューの組み合わせは最適な物に近づいていく。

【0046】

再評価の結果、レベル3の付いたV○QにGrantを与え83-1、Grantを得たV○Qと同じ入出力にあるV○Qの送出権を削除83-2する。

【0047】

続いて、レベル2のV○Q85にGrantを与え85-1、Grantを得たV○Qと同じ入出力にあるV○Qの送出権を削除85-2する。

【0048】

続いて、レベル1のV○Q86にGrantを与え86-1、Grantを得たV○Qと同じ入出力にあるV○Qの送出権を削除86-2する。

【0049】

最後に、レベル0のV○QにもGrantを与える87。これで、全入力、出力の組み合わせにGrantを与え終り、アービトレーションの一工程は終了する。

【0050】

図12は、図11のフローチャートに従い行った、 4×4 スイッチのアービトレーションを示す図である。マトリクス21は、アービタ14の全V○Qレベル収集部121に集められた、ARB-REQ生成部13でレベル付けされた全V○QのV○Qレベルを視覚的に分かりやすく表したものである。マトリクスの行は入力回線番号を表しており、列は出力回線番号を表している。例えば、マトリクス21において、入力回線番号3で、出力回線番号1のV○Qレベルは10であることがこのマトリクスから分かる。

【0051】

入力回線毎、及び出力回線毎に上述の勝ち抜き戦が行われ、その後、図10を用いて説明した再評価が行われる。マトリクス22は、その結果を示す。

【0052】

マトリクス23は、マトリクス22の中からレベル3のV○Qを選び出し、それにGrantを与えたことを示している。Grantが与えられたV○Qと同じ入力回線番号又は出力回線番号に属するV○Qに対しては、同一アービトレーション周期において、Grantを与えることが出来ないので、それらから送出権を剥奪する。マトリクス24中の“×”は、送出権を剥奪したことを示している。

【0053】

次に、まだGrantの出でていないV○Qのレベルを、ARB-REQ生成部で作られたV○Qレベルに変換し直す。マトリクス25は、その変換後のマトリクスを示している。再び勝ち抜き戦、レベルの再評価を行なう。マトリクス26は、その結果を示している。マトリクス26において、レベル3のV○Qに対してGrantが与えられる。ここでは、入力回線番号2で出力回線番号3のV○Qに送出権が与えられる。マトリクス27は、それに送出権が与えられたことを

示している。Grantが与えられたV○Qと同じ入力回線番号、出力回線番号に属するV○Qから、送出権を剥奪する。マトリクス28は、それから送出権を剥奪したことを示している。次にレベル2のV○Qに送出権を与える。マトリクス29は、それに送出権を与えたことを示している。図6で説明した例では、上記工程により、すべての入出力の組み合わせにGrantを与えたので一連のアービトレーション処理が終了する。これ以外の例では、レベル1及びレベル0のV○QにGrantを与える処理が必要になる場合もある。この場合に、図11に示すフローに従い、Grantを与える処理が行われる。

【0054】

図13は、本発明のパケット処理装置の他の実施例を示す。本実施例では、別線ではなくセグメントのヘッダ部分にARB-REQ情報を付与しインチャネルでこれをアービタに送る。

【0055】

図1に示したパケット処理装置との違いは、信号線118により、一旦、ARB-REQ情報をARB-REQ付与部111に送り、セグメント113のヘッダ部分にARB-REQ情報114を付与してアービタ14に送る方式を採用している点、およびアービタ14がクロスバスイッチ19内に内蔵されている点である。図1の実施例と同様、ARB-REQ情報はアービタ14に集められ、ARB-REQ情報からどのV○QにGrantを与えるかを決定する。その後、Grant情報はARB-ACKとして、信号線129により、アービタからスイッチング済みセグメント115のヘッダー116に付与され。Grant情報は、ARB-ACK回収部112において集められて、信号線119により、V○Q制御部12に送られる。V○Q制御部12は、どのV○Qからセグメントを送出するかを10の入力バッファに指示する。

【0056】

この方式の利点は、アービタ用に信号線を用意する必要がないため、信号線数を削減することができ、ハードウェアの作りをシンプルに出来る点である。

【0057】

図14と図15は、本発明のキュー情報管理アービタと、キューのパケットの

有無のみを見てラウンドロビンを行うアービトレーション方式(若山浩二 他, “大容量パケットスイッチ構成法に関する一検討” 信学技報 IN98-102 参照)と同じ条件の下で、シミュレーションを行なった時の結果である。シミュレーションの条件は入力回線数4 出力回線数4 の 4×4 の入出力バッファ型クロスバススイッチを想定している。

【0058】

図14はユニフォームトラヒックをかけた時の、従来方式アービタと本願において提案される方式のアービタとの平均遅延時間の分布グラフである。縦軸92にはDelay (Segment), 横軸91にはLoad Rate (%)をとっている。回線負荷が高くなるにつれ従来方式アービタの遅延量は大きくなるのに対し、提案方式アービタでは遅延時間の増加を抑えることが出来ている。

【0059】

ユニフォームトラヒックであっても、回線負荷が高くなると、多少トラヒック状態に偏りが生じてしまい、そのためキュー長、パケット送出間隔を見ている提案方式アービタの方がパケットの有無のみを見ている従来方式アービタよりも遅延時間を抑えられる。

【0060】

図15は4入力回線のうち、一つの入力回線に他よりも負荷の高いトラヒックをかけた時の高負荷キューと、低負荷キューの遅延分布をシミュレーションした時の結果である。縦軸102にProbability (Delay time > d), 横軸101にDelay (segment) を取っている。

【0061】

菱形のプロットが高負荷のトラヒックが掛かっている入力ポートの遅延分布で、三角のプロットが低負荷のトラヒックが掛かっている入力ポートの遅延分布をあらわしている。

【0062】

高負荷キューの遅延分布を見ると、提案方式アービタは従来方式アービタに比べ、遅延時間をかなり抑えることが出来ていることが分かる。高負荷の掛かっているV〇Qのキュー長は他のV〇Qに比べ大きくなっている。キュー長を見てア

ービトレーションを行なう提案方式では、高負荷キューに多くGrantを出すことが出来るため、高負荷キューの遅延時間を大きく抑えることが出来る。

【0063】

一方その影響で低負荷キューにインパクトを与えることになるが、本発明では、送出間隔も見ているため、低負荷キューに与える影響も抑えることが出来る。

【0064】

【発明の効果】

キュー長を管理しているため、不均衡負荷が掛かった時でも効率良いスイッチングを行なうことを可能とし、高負荷キューの遅延時間を効果的に抑え、かつ低負荷キューにもその影響をほとんど与えることなくパケットを送出することを可能とする。

低負荷キューでもユーザが設定したタイムアウト時間になるとV○Q送出要請リクエストは最大レベルを取ることになる。このため、次処理の勝ち抜き戦で、入力方向、出力方向でともに勝ち残りやすくなり、レベルの再評価を行なった際にも最高レベルになる可能性が高い。よって、そのV○Qはセグメントの送出権を獲得しやすくなる。従って低負荷キューのstarvationを防ぐことが出来る。

【0065】

本発明を用いることで、均衡負荷にも不均衡負荷にも対応可能なアビタを提供することが出来る。

【図面の簡単な説明】

【図1】

本発明のパケット処理装置の一実施例を説明するためのブロック図である。

【図2】

図1のV○Q制御部12とARB-REQ生成部13の構成例を示すブロック図である。

【図3】

図1のARB-REQ生成部でV○Qにレベルを割り振るレベル付与マトリクスの一例である。

【図4】

図1のARB-REQ生成部でV○Qにレベルを割り振るレベル付与マトリクスの一例（送出間隔重視）である。

【図5】

図1のARB-REQ生成部でV○Qにレベルを割り振るレベル付与マトリクスの一例（キュー長重視）である。

【図6】

図1のアービタ14の一構成例を示すブロック図である。

【図7】

本発明におけるV○Qレベルマトリクスの概念を説明するための図である。

【図8】

本発明における勝ち抜き戦の概念を説明するための図である。

【図9】

本発明の入力別、出力別勝ち抜き戦の勝敗全組み合わせを表す図である。

【図10】

本発明のレベル再評価表の概念を説明するための図である。

【図11】

図1のアービタ14が実行するアルゴリズムを説明するためのフローチャートである。

【図12】

図1のアービタ14の処理を説明するための図である。

【図13】

本発明のパケット処理装置の他の実施例の全体構成を表すブロック図である。

【図14】

本発明によるアービタと、従来方式アービタの99%遅延を示すグラフである

【図15】

不均衡負荷を加えた場合の、本発明によるアービタと従来方式アービタのキュー長分布を示すグラフである。

【図16】

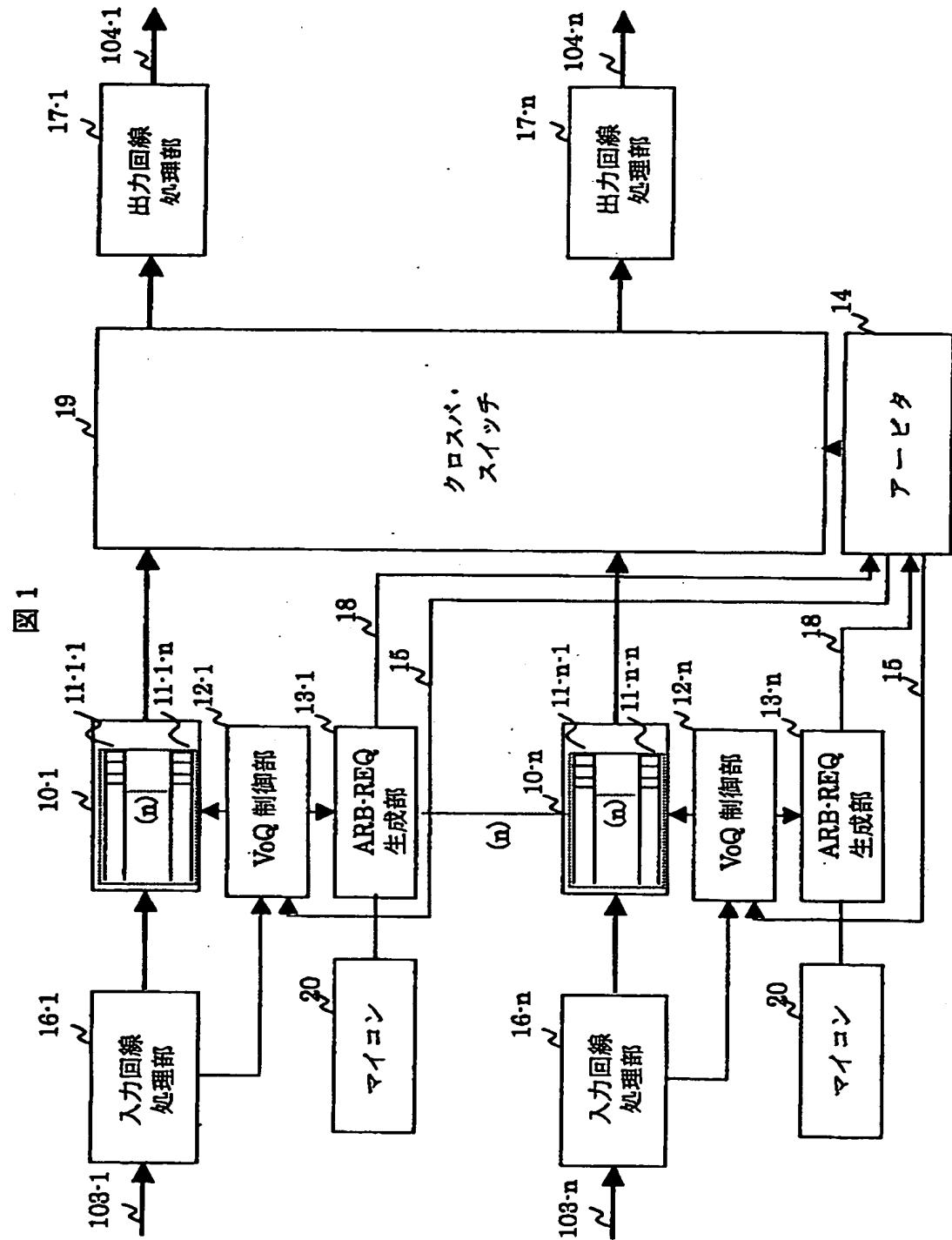
高負荷キューのオーバーフロー、低負荷キューのstarvationの概念を説明するための図である。

【符号の説明】

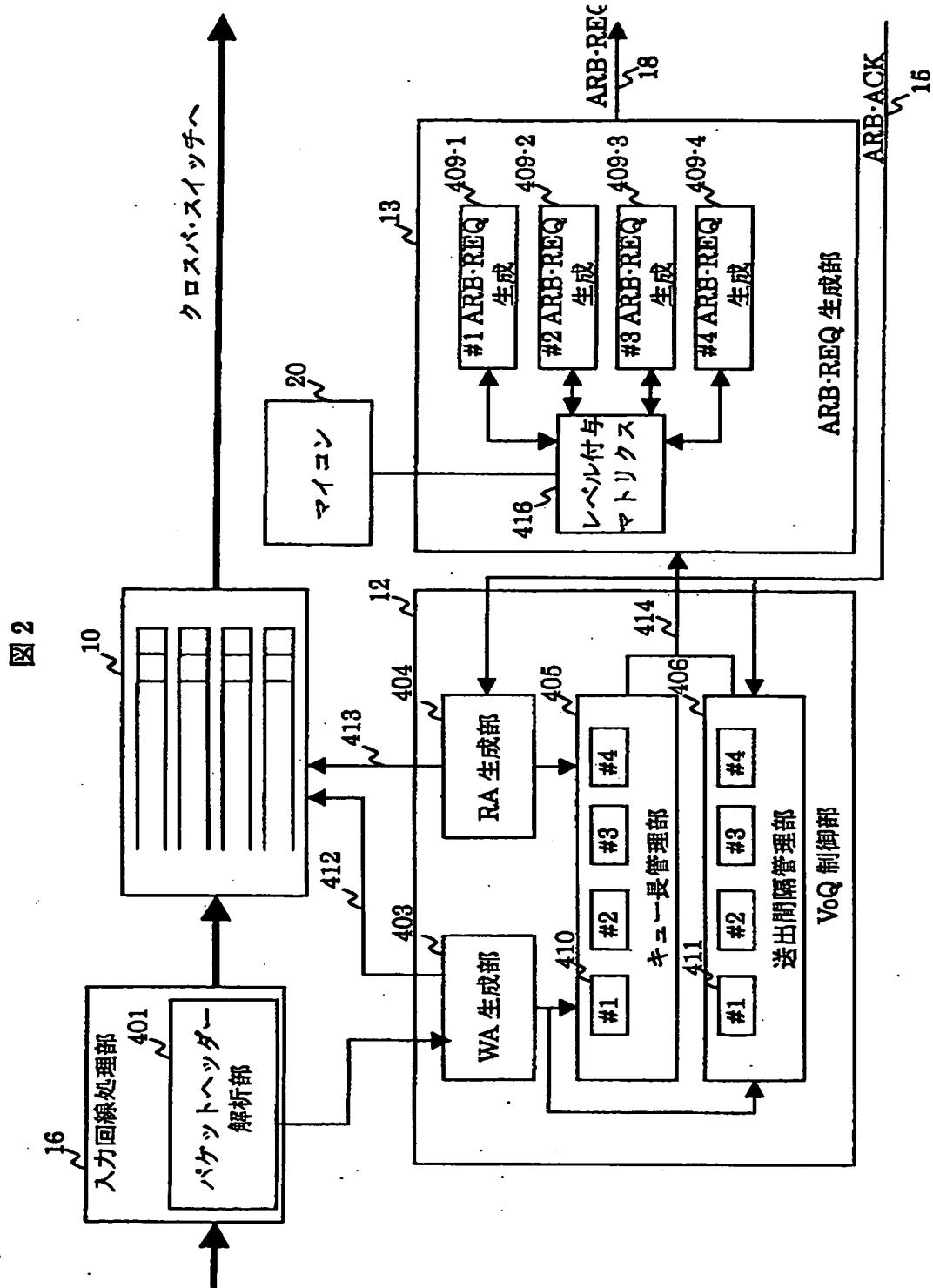
10…入力バッファ、11…V○Q、12…V○Q制御部、13…ARB-REQ生成部、14…アービタ、15…ARB-ACK信号線、19…クロスバスイッチ。

【書類名】 図面

【図1】



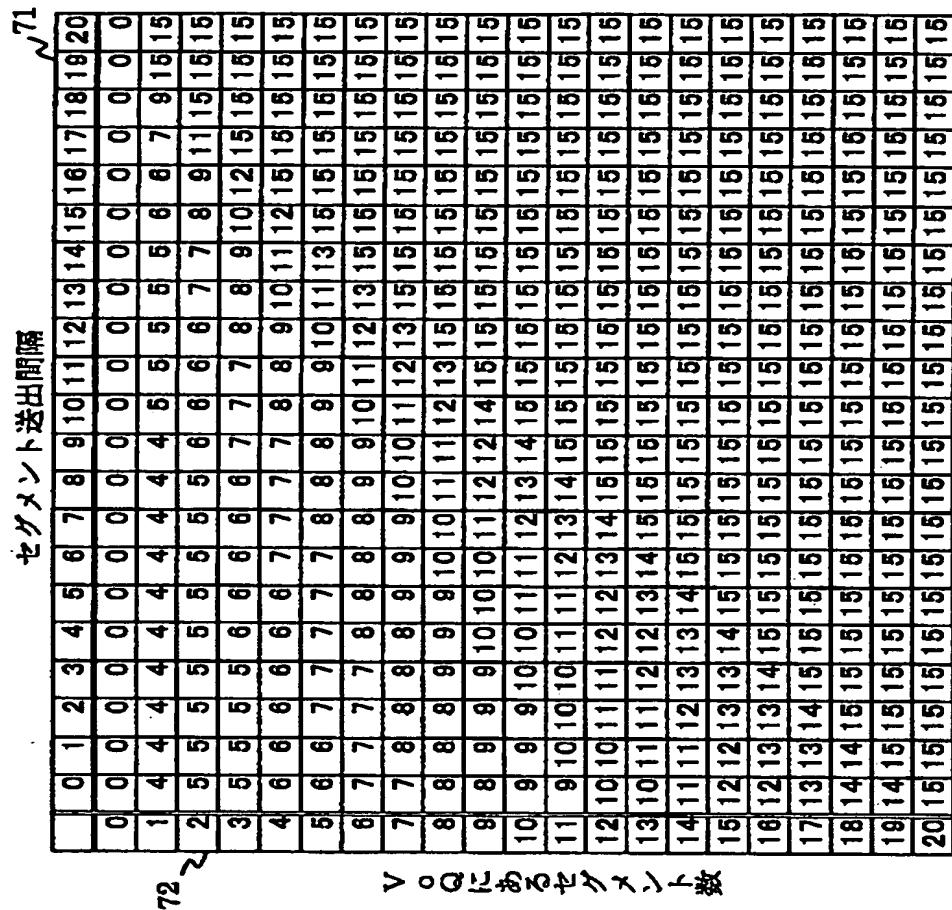
【図2】



【図3】

23

VoQ レベル付与マトリクス



【図4】

図4
VoQ レベル付与マトリクス

		セグメント送出間隔																		
		71																		
		72																		
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	4	4	4	4	5	5	6	6	9	15	15	16	15	15	15	15	15	15	15	15
2	5	5	5	5	6	6	7	9	15	15	15	15	15	15	15	15	15	15	15	15
3	5	5	6	6	7	8	9	12	15	15	15	15	15	15	15	15	15	15	15	15
4	6	6	7	7	8	9	11	15	15	15	15	15	15	15	15	15	15	15	15	15
5	6	7	7	8	9	10	13	16	15	15	15	15	15	15	15	15	15	15	15	15
6	7	7	8	8	9	10	12	15	15	15	15	15	15	15	15	15	15	15	15	15
7	7	8	8	9	10	11	13	15	15	15	15	15	15	15	15	15	15	15	15	15
8	8	8	8	9	10	11	12	15	15	15	15	15	15	15	15	15	15	15	15	15
9	8	9	10	10	12	14	15	15	15	15	15	15	15	15	15	15	15	15	15	15
10	9	9	10	11	13	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
11	9	10	11	12	14	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
12	10	11	12	13	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
13	10	11	12	14	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
14	11	12	13	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
15	12	13	14	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
16	12	13	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
17	13	14	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
18	14	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
19	14	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
20	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15

V_oQ にあるセグメント数

【図5】

図5

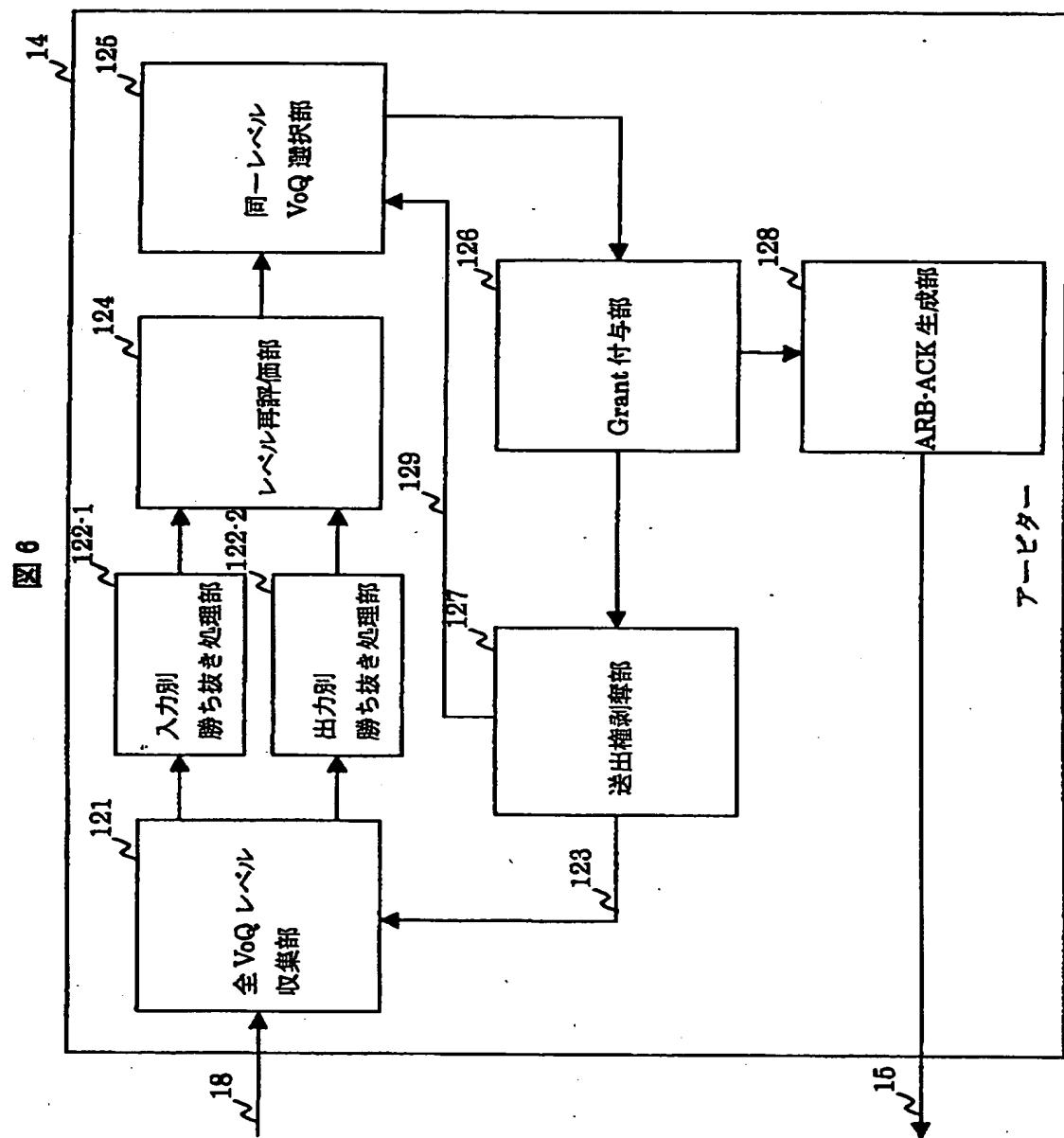
VoQ レベル付与マトリクス

	セグメント送出間隔																			
	71																			
0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
3	2	6	6	6	6	6	6	7	7	7	7	7	7	8	8	9	10	11	12	13
4	3	7	7	7	8	8	8	9	9	9	10	11	11	12	13	15	15	15	15	15
5	4	8	8	8	9	9	9	9	9	9	10	10	11	11	12	13	14	15	15	15
6	5	9	9	9	9	9	9	9	9	9	10	10	11	11	12	13	14	15	15	15
7	6	10	10	11	11	12	12	13	14	14	15	15	15	15	15	15	15	15	15	15
8	7	11	11	12	12	13	13	14	14	15	15	15	15	15	15	15	15	15	15	15
9	8	12	12	13	13	13	14	14	15	15	15	15	15	15	15	15	15	15	15	15
10	9	14	14	14	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
11	10	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
12	11	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
13	12	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
14	13	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
15	14	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
16	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
17	16	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
18	17	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
19	18	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
20	19	16	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15

VoQにおけるセグメント数

72

【図6】



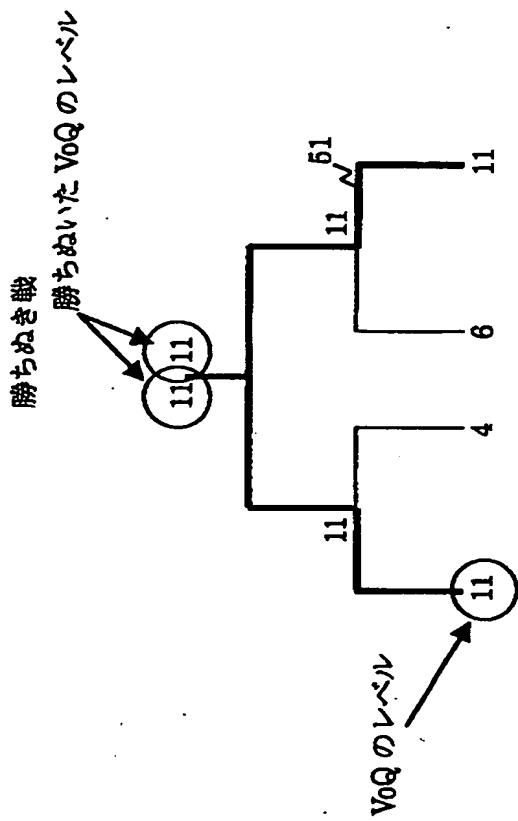
【図7】

図7

		出力回線番号			
		1	2	3	4
132		1			
入力回線番号	1				
2			133		133
3					
4			133		

【図8】

8



【図9】

図9

141

入力方向	出力方向
单一勝ち残り	单一勝ち残り
单一勝ち残り	複数勝ち残り
複数勝ち残り	单一勝ち残り
单一勝ち残り	負け
複数勝ち残り	負け
負け	单一勝ち残り
負け	複数勝ち残り
負け	負け

【図10】

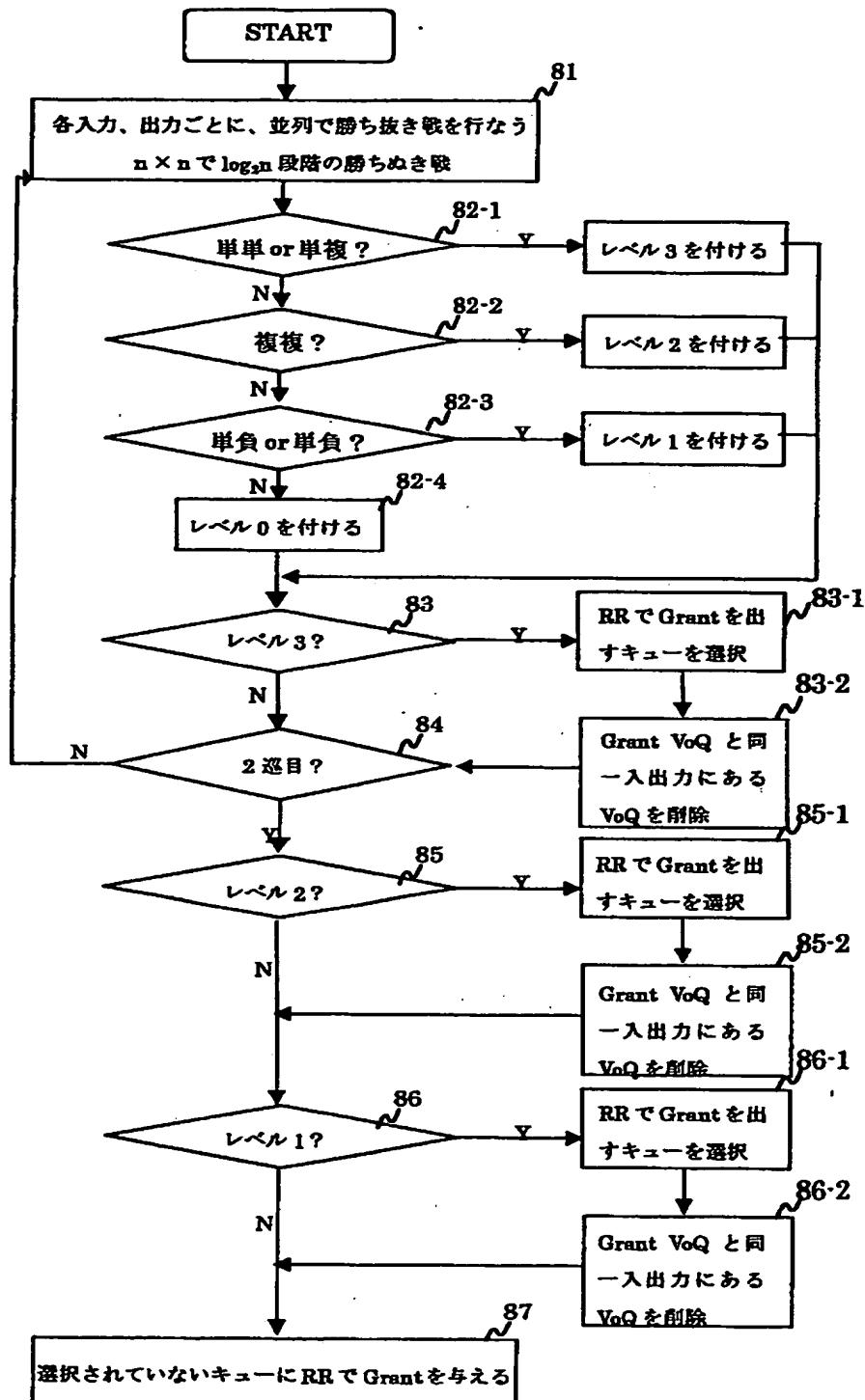
図10

61

レベル再評価表			
	単勝	複勝	負け
単勝	レベル8	レベル3	レベル1
複勝	レベル3	レベル2	レベル1
負け	レベル1	レベル1	レベル0

【図11】

図11



【図12】

21

出力回線			
1	2	3	4
1	2	3	4
1	12	8	0
2	0	4	6
3	10	4	7
4	0	0	15

22

出力回線			
1	2	3	4
1	8	0	5
2	0	4	6
3	10	4	7
4	0	0	15

23

出力回線			
1	2	3	4
1	8	0	5
2	0	4	6
3	10	4	7
4	0	0	15

24

出力回線			
1	2	3	4
1	8	0	5
2	0	4	6
3	10	4	7
4	0	0	15

25

出力回線			
1	2	3	4
1	8	0	5
2	0	4	6
3	10	4	7
4	0	0	15

26

出力回線			
1	2	3	4
1	8	0	5
2	0	4	6
3	10	4	7
4	0	0	15

27

出力回線			
1	2	3	4
1	8	0	5
2	0	4	6
3	10	4	7
4	0	0	15

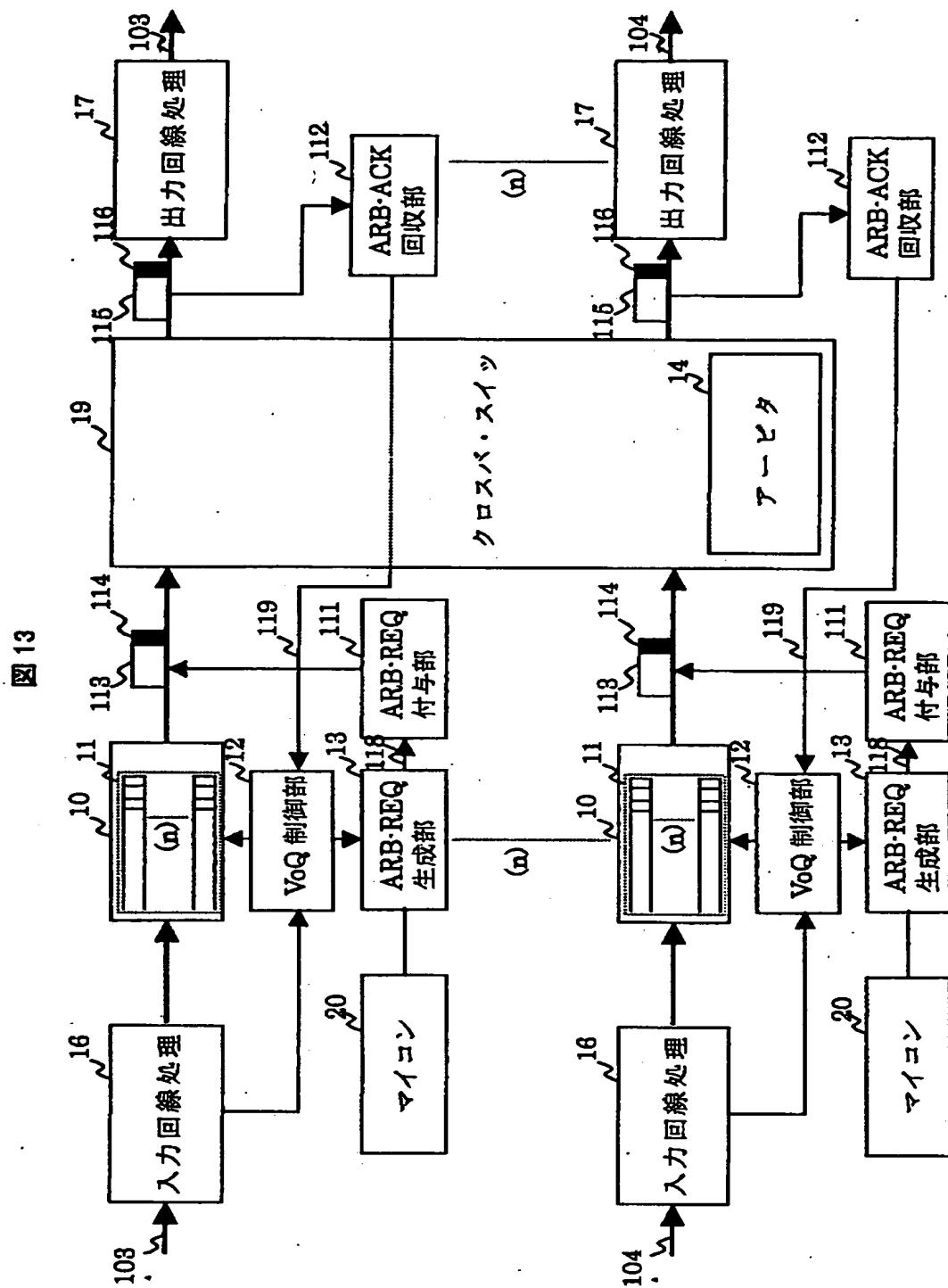
28

出力回線			
1	2	3	4
1	8	0	5
2	0	4	6
3	10	4	7
4	0	0	15

29

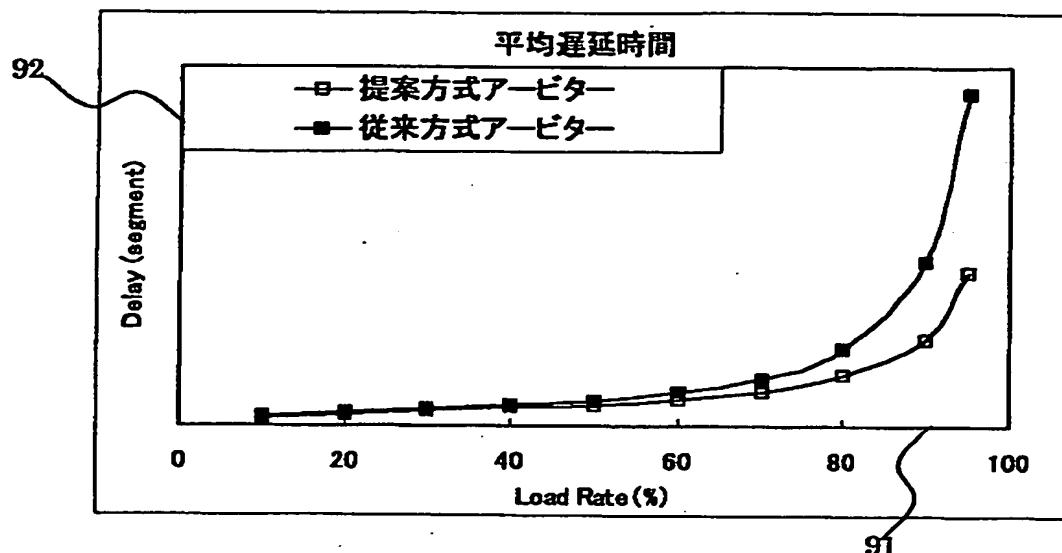
出力回線			
1	2	3	4
1	8	0	5
2	0	4	6
3	10	4	7
4	0	0	15

【図13】



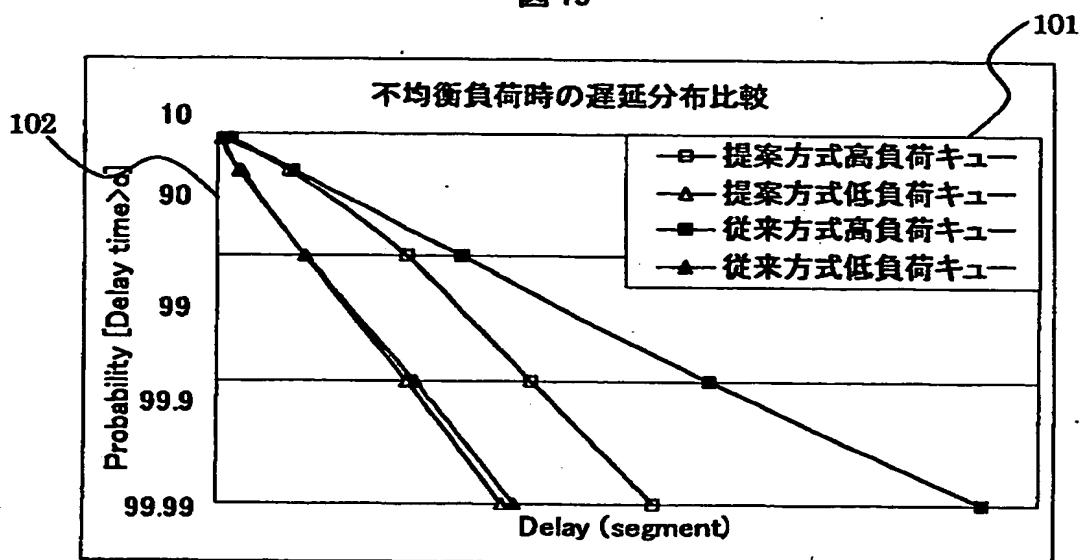
【図14】

図14

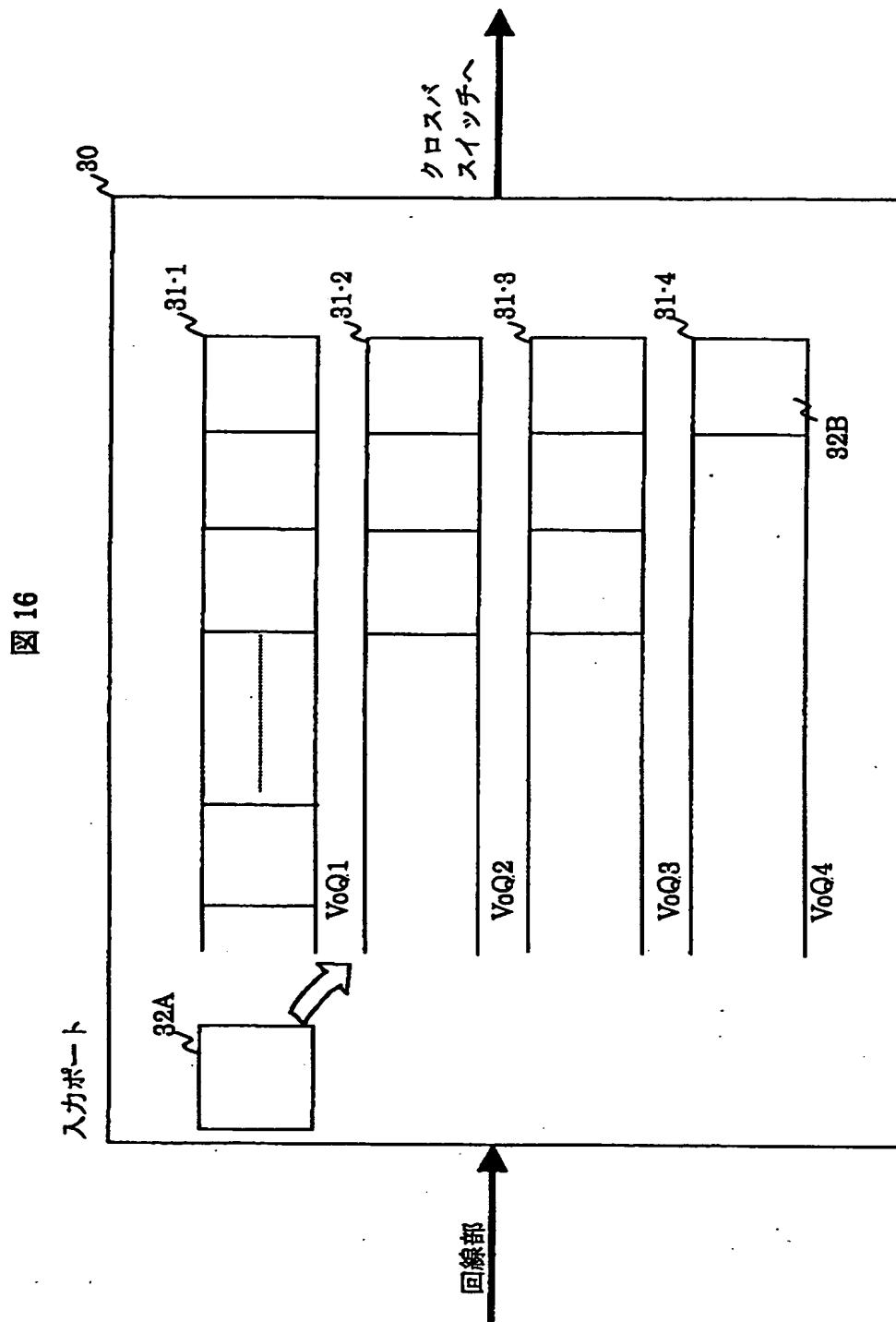


【図15】

図15



【図16】



【書類名】 要約書

【要約】

【課題】 不均衡負荷下においても高負荷キューの遅延時間を抑えることにより、バッファのオーバーフローを抑制し、一方、低負荷キューからも高負荷キューの影響を与えることなくパケットを送出することができるアービトレーション方式を提案する。

【解決手段】 V○Qのパケット送出間隔と、キュー長に基づき、クロスバスイッチに対するパケットの送出権を何れのV○Qに与えるかを決定する。

【選択図】 図1

認定・付加情報

特許出願の番号 特願2001-047548
受付番号 50100252089
書類名 特許願
担当官 第八担当上席 0097
作成日 平成13年 2月26日

＜認定情報・付加情報＞

【提出日】 平成13年 2月23日

次頁無

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地

氏 名 株式会社日立製作所